

Evolving Fuzzy Image Segmentation with Self-Configuration

A. Othman¹, H.R. Tizhoosh², F. Khalvati³

¹ Dept. of Information Systems, Computers & Informatics, Suez Canal University, Egypt :: a.othman@ci.suez.edu.eg

² Centre for Pattern Analysis and Machine Intelligence, University of Waterloo, Canada :: tizhoosh@uwaterloo.ca

³ Sunnybrook Health Sciences Centre, University of Toronto, Canada :: farzad.khalvati@sri.utoronto.ca

Abstract—Current image segmentation techniques usually require that the user tune several parameters in order to obtain maximum segmentation accuracy, a computationally inefficient approach, especially when a large number of images must be processed sequentially in daily practice. The use of evolving fuzzy systems for designing a method that automatically adjusts parameters to segment medical images according to the quality expectation of expert users has been proposed recently (Evolving fuzzy image segmentation – EFIS). However, EFIS suffers from a few limitations when used in practice mainly due to some fixed parameters. For instance, EFIS depends on auto-detection of the object of interest for feature calculation, a task that is highly application-dependent. This shortcoming limits the applicability of EFIS, which was proposed with the ultimate goal of offering a generic but adjustable segmentation scheme. In this paper, a new version of EFIS is proposed to overcome these limitations. The new EFIS, called self-configuring EFIS (SC-EFIS), uses available training data to self-estimate the parameters that are fixed in EFIS. As well, the proposed SC-EFIS relies on a feature selection process that does not require auto-detection of an ROI. The proposed SC-EFIS was evaluated using the same segmentation algorithms and the same dataset as for EFIS. The results show that SC-EFIS can provide the same results as EFIS but with a higher level of automation.

I. INTRODUCTION

Evolving fuzzy image segmentation (short EFIS [19]) has been recently introduced to solve the parameter setting problem (e.g., fine-tuning) of different segmentation techniques. EFIS has been designed with emphasis on acquiring and integrating user feedback into the fine-tuning process. As a result, EFIS is suitable for all applications, such as medical image analysis, in which an experienced and knowledgeable user provides evaluative feedback of some sort with respect to the quality, i.e., accuracy, of the image segmentation.

Image segmentation is the grouping of pixels to form meaningful clusters of pixels that constitute objects (e.g., organs, tumours), a task with various applications in medical image analysis including measurement, detection, and diagnosis. Image segmentation can be roughly categorized into two main classes of algorithms; non-parametric-based (e.g., atlas-based segmentation) and parametric-based (e.g., thresholding, region growing) algorithms. The former is based on a model which usually does not require parameters whereas the latter is based on some parameters that must be adjusted in order to obtain reasonable segmentation results. Parameter-based segmentation algorithms always face the challenge of parameter adjustment; a parameter tuned for a particular set of images may perform poorly for a different image category.

On the other hand, in a clinical setting such as in a hospital, the final outcome of image segmentation algorithms usually need to be modified (i.e., manually edited) and approved by a an expert (e.g., radiologist, oncologist, pathologist). The clinical ramifications of not verifying the correctness of segments include missing a target (resulting in a less effective therapy) or increased toxicity if the target is over-segmented. The frequent expert intervention to correct the results, in fact, generates valuable feedback for a learning scheme to automatically adjust the segmentation parameters.

EFIS is an images segmentation scheme that evolves fuzzy rules to tune the parameters of a given segmentation algorithm by incorporating the user feedback which is provided to the system as corrected or manually created segmentation results called *gold standard images*. EFIS represents a new understanding of how image segmentation should be designed in the context of observer-oriented applications. Naturally, EFIS needs to be further improved and extended in order to exploit the full potential of its underlying evolving mechanism in relation to the user feedback. The original design of EFIS as presented in [19] requires pre-configurations of a few steps which should be set for a given image set and the segmentation algorithm to which EFIS is integrated. This limits the efficiency of EFIS; either the algorithm should be pre-configured for each dataset and/or segmentation algorithm or it is possible that a fixed pre-configuration will adversely affect its performance. In this paper, we present a new and extended version of EFIS which we call self-configuring EFIS (short SC-EFIS) that has a higher level of automation. The new extension of EFIS proposed in this paper will enhance EFIS through removing these limitations by introducing self-configuration into different stages of EFIS.

This paper is organized as follow: In section II, a brief review of the EFIS (evolving fuzzy image segmentation) will be provided. In section III, we critically point to the shortcomings of EFIS. The section IV reviews the literature on feature selection as this is the major improvement in SC-EFIS compared to EFIS. In section V, we present the proposed self-configuring EFIS (SC-EFIS). In section VI, experiments are described and the results are presented and analyzed. Finally, section VII concludes the paper.

II. A BRIEF REVIEW OF EFIS

The concept of Evolving Fuzzy Image Segmentation, EFIS, was proposed recently [19]. The problem that EFIS attempts to address is parameter adjustment in image segmentation. The basic idea of EFIS is to adjust the parameters of segmentation

Algorithm 1 EFIS [19]: Simplified Overview

Training: Stage 1

Determine the parent algorithms and their parameters
Read the training images and their gold standard images
Via exhaustive/trial-and-error comparisons with gold standard images, determine the best segments and the best parameter(s) that generate the best segments

Training: Stage 2

Read the available training images
Determine regions of interest (ROIs) around each segment

Save ROIs for each image

Training: Stage 3

Set the number of seeds inside the segments, and the number of rules to be extracted

for all images **do**

for all seeds **do**

 Determine a new seed point inside the ROI

 Extract features from the seed point's neighbourhood

 Save features and best parameters in matrix M

end for

end for

Generate fuzzy rules from the rule matrix M

Save the rule matrix M and the generated rules

Online: Evolving Phase

Load the fuzzy rules and the rule matrix M

Read a new image

Detect ROI

Determine seed points inside ROI

Extract features from the seed point's neighbourhood

Perform fuzzy inference to generate output(s):
parameters = FUZZY-INFERENCE(RULES)

Apply the parameters to segment the image

Display the segment and wait for the user feedback (user generates a gold standard image by editing the segment)

Rule Evolution - Invisible to User

Determine the best output(s) (via comparison of segments with the gold standard image)

if (Pruning) the features/parameters not seen yet **then**

 Add new rows to the rule matrix

 Generate fuzzy rules from the rule matrix M

 Save the rule matrix M and the generated rules

end if

to increase the accuracy by using user feedback in form of corrected segments. To do so, EFIS extracts features from a region inside the image and assigns them to the best parameter exhaustively detected. Clustering or other methods are then used to generate fuzzy rules, which are then continuously updated when new images are processed. The simplified pseudo-code of EFIS is given in Algorithm 1.

EFIS needs to be trained for specific algorithms and image categories [19]. In other words, in order to employ EFIS, the following components must be pre-designated:

- Parent algorithm: any segmentation algorithm with at least one parameter that affects its accuracy (e.g., global thresholding, statistical region merging)
- Parameter(s) to be adjusted (e.g., thresholds, scales)

- Images and corresponding gold standard images
- Procedure to find optimal parameters (e.g., brute force or trial-and-error via comparison with the gold standard images)

Once the above-mentioned components are available/defined, the following steps need to be specified:

- ROI-detection algorithm: An algorithm that detects the region of interest (ROI) around the subject to be segmented by EFIS.
- Procedure for feature extraction around available seed points: Methods like SIFT are used to generate seed points. But a certain number of expressive features should be calculated in the vicinity of each seed point to be fed to fuzzy inference system.
- Rule pruning: Upon processing a new image, a new rule can be learned only if the features and corresponding output parameters had not been observed previously. In other words, by looking at the difference between an input (features plus outputs) with all rules in the database, the information of a new image is added only if not captured by existing rules.
- Label fusion: When EFIS is used with multiple algorithms at once, the segmentation results are fused using a fusion method namely STAPLE algorithm [26].

EFIS includes two main phases namely training and testing. In training phase, images with their gold standard results are fed to the algorithm where features are extracted from each image. The parent algorithm, e.g., thresholding, is applied to each image and the results are compared to the gold standard image. The algorithm's parameters are continuously changed until the best possible result is achieved. The best parameter which yields the best result (i.e., the highest agreement with the gold standard image) along with the image feature extracted in the previous stage are stored. Once all training images are processed, the fuzzy rules are generated from the stored data using a clustering algorithm.

In testing phase, new images are first processed to extract features. Next, the image features are fed to the fuzzy inference system to approximate the parameters. The parent algorithm is then applied to the input image using the estimated parameter. EFIS can address both single-parametric and multi-parametric problems. EFIS was applied to three different thresholding algorithms and significant improvements in terms of segmentation accuracy were achieved [19].

III. CRITICAL ANALYSIS OF EFIS

Although EFIS has demonstrated to improve the segmentation results [19], some of its underlying steps may limit its applicability mainly because these steps have been designed in an ad-hoc fashion and tailored to the specific test images and algorithms namely breast ultrasound and thresholding. In this section, we examine the limitations of EFIS and lay out how they should be addressed via self-configuration.

EFIS calculates the features inside a rectangle that constitutes the region of interest, ROI. Within this region, n feature are calculated using scale-invariant feature transform

(SIFT) [13], [14]. In designing the ROI-detection algorithm, it is assumed that the ROI will be dark based on the characteristics of test images used (breast lesions in ultrasound are hypoechoic, meaning they are darker than surrounding tissue). This means that EFIS needs a detection algorithm for any new image category (application) to correctly recognize the region of the image containing the object of interest. In addition, similar to any other detection algorithm, if it fails, then EFIS will not be able to perform. We will remove this dependency by redesigning the feature extraction stage.

In order to calculate features within the ROI, EFIS uses a fixed number of landmarks, called seed points, which are delivered by SIFT. These n fixed key points, with $n = 10$, is set for all images regardless of their content. Of course, an arbitrary number of features may not be able to characterize all types of images. We will eliminate this limitation of EFIS by automatically setting the number of seed points for different image categories.

EFIS constructs a fixed sized window of 40×40 pixels around each landmark (seed point) to calculate the features. A self-configuring EFIS has to automatically set the window size during a pre-processing stage in order to optimally define the feature neighbourhood.

EFIS uses a fixed number of manually selected features, namely 18 features which proved to perform well on the breast ultrasound images. It is intuitively clear that this may not be a flexible approach to capture the image content. Any set of images with some common characteristics may need a different set of features for the evolving fuzzy systems to effectively estimate the parameters of the segmentation.

In the proposed extension of EFIS algorithm, we will address these shortcomings by introducing a pre-processing (self-configuration) stage where the settings are undertaken automatically. As apparent from the list above, feature selection seems to be the core of EFIS lack of automation. In following section, therefore, we will briefly review feature selection methods.

IV. FEATURE SELECTION

Providing relevant features to a learning system will increase its ability to generalize and hence elevate its performance. Feature selection is the process of selecting the most relevant features out of a larger group of features so that either redundant or irrelevant features are removed. Redundant features add no new information to the system, and irrelevant features may confuse the system and decrease its ability to learn efficiently. Feature selection may be conducted according to one of four schemes [17]:

- **Filter feature selection** methods work directly on the available data and select features based on the data properties. They are independent of any learning methods [21], [12], [1].
- **Wrapper feature selection** methods may evaluate features but without consideration of the structure of the classifier [12].
- **Embedded feature selection** treats the learning and feature selection aspects as one process.

- **Hybrid systems** may combine wrapper and filter approaches [3].

Feature selection may also be categorized into three main branches: supervised, semi-supervised, and unsupervised.

A. Supervised Feature Selection

In supervised feature extraction, the selection of a set of features from a larger number of features is based on one of three characteristics [17]: 1) features of a size that optimize an evaluation measure, 2) features satisfying a condition in the evaluation measure, and 3) features that best match a size and evaluation measure. Supervised feature selection methods deal primarily with the classification problems, in which the class labels are known in advance [20]. Numerous studies have investigated supervised feature selection using the measures of the information theoretic [15] and Hilbert-Schmidt independence criterion [22].

B. Semi-Supervised Feature Selection

The concept of semi-supervised feature selection has emerged recently as a means of addressing situations in which insufficient labels are available to cover the entire training data [27] or in which a substantial portion of the data are unlabelled. Traditional supervised feature selection techniques are generally ineffective under such circumstance. Semi-supervised feature selection is therefore employed for the selection of features when not enough labels are available. A semi-supervised feature selection constraint score that takes into account the unlabelled data has been proposed in [10]. The literature also contains proposals for numerous semi-supervised techniques based on spectral analysis [27], a Bayesian network [5], a combination of a traditional technique with feature importance measure [2], or the use of a Laplacian score [6]. Although semi-supervised selection does not require a complete set of class labels, it does need some.

C. Unsupervised Feature Selection

Unsupervised feature selection is the process of selecting the most relevant non-redundant features from a larger number of features without the use of class labels. Mitra et al. [16] proposed an unsupervised feature selection algorithm based on feature similarity. They used a maximum information compression index to measure the similarities between features so that similar features could be discarded. He et al. [8] proposed an unsupervised feature selection technique that relies on the Laplacian score to indicate the significance of the features. Zhao et al. [28] used spectral graph theory to develop a new algorithm that unifies both supervised and unsupervised feature selection in one algorithm. They applied the spectrum of the graph that contains the information about the structure of the graph in order to measure the relevance of the features. Cai et al. [4] proposed a new unsupervised feature selection algorithm called Multi-Cluster Feature Selection, in which the features selected are those that maintain the multi-cluster structure of the data. Farahat et al. [7] present a novel unsupervised greedy feature selection algorithm consisting of two parts: a recursive technique for calculating the reconstruction error of the matrix of features selected, and a greedy algorithm for feature selection. The method was tested on six different

benchmark data sets, and the results show an improvement over state-of-the-art unsupervised feature selection techniques.

D. Features for SC-EFIS

In order to eliminate the major shortcomings of EFIS with respect to inflexible and static feature selection, and in order to not assume availability of class labels, we chose unsupervised feature selection, specially the previously mentioned five popular unsupervised feature selection algorithms to characterize images for training the evolving fuzzy system. These five methods, along with an additional correlation-based method, were combined to produce an ensemble of final relevant features that could be used for training.

In the remaining of the paper, the output matrices of these techniques are denoted as follows:

- Mitra et al. [16]- F_F (feature similarity).
- He et al. [8]- F_L (Laplacian score).
- Zhao et al. [28]- F_P (spectral graph).
- Cai et al. [4]- F_M (multi-cluster).
- Farahat et al. [7]- F_G (greedy algorithm).
- F_C (correlation method).

V. SELF-CONFIGURING EFIS (SC-EFIS)

This section introduces a new version of EFIS, namely a self-configuring evolving fuzzy image segmentation (SC-EFIS) which represents a higher level of automation compared to the original EFIS scheme. The proposed SC-EFIS scheme consists of three phases; self-configuration phase, training phase, and online or evolving phase. In the following, each of these phases are described in detail.

A. Self-Configuring Phase

In the self-configuring phase (Algorithm 2), all available images are processed in order to determine two crucial factors: 1) the size of the feature area around each seed point, and 2) the final features to be used for the current image category.

The $Z \times Z$ rectangle around each SIFT point to be used for feature calculation is determined based on different sizes of all available images (algorithm 2). Following this step, the set of features that should be used for the available images is selected from a large number of features which are calculated for each image from the vicinity of the SIFT points located in the entire image (since there is no longer an ROI) (Fig. 1). This process starts with the determination of the number of SIFT points N_F that should be used in the current image (algorithm 2). This step is identical to the procedure used in the EFIS training phase, as previously explained in section II, with three exceptions: the SIFT points are detected across the entire image (as opposed to selecting SIFT points inside an ROI as a subset of the image), the final number N_F of SIFT seed points is not fixed, and the points returned are separated from each other by Z in each direction. For all N_F seed points, features are extracted from a rectangle R_C around each point, based on the discrete cosine transform (D_C) of R_C , the gradient magnitude (G_M) of R_C , the approximation coefficient matrix

A_C of R_C (computed using the wavelet decomposition of R_C), and the SIFT descriptors D_S . The following set of features is extracted (Algorithm 2):

- 1) The mean, median, standard deviation, co-variance, mode, range, minimum, and maximum of R_C , D_{R_C} , and A_{R_C} , and $G_{M_{R_C}}$ (32 features)
- 2) The mean, median, standard deviation, co-variance, range, minimum, maximum, and zero population of D_S (eight features) with the minimum of D_S changed to be the minimum number after zero
- 3) The contrast, correlation, energy, and homogeneity of the gray level co-occurrence matrices (computed in four directions 0° , 45° , 90° , and 135°) of R_C , D_{R_C} , and A_{R_C} , and $G_{M_{R_C}}$ (64 features)
- 4) The contrast, correlation, energy, and homogeneity of the gray level co-occurrence matrices (computed in only one directions of 0°) of D_S (four features)
- 5) A feature matrix F_1 of size $N_F \times N_T$ generated for I (in this case $N_T = 108$)

Algorithm 2 Self-Configuration Phase

- 1: Set the variables and initialize all matrices
 - 2: Read the available images I_1, I_2, \dots, I_{N_I} .
 - 3: Read the size of the images, namely all rows R_1, R_2, \dots, R_{N_I} , and all columns C_1, C_2, \dots, C_{N_I} .
 - 4: Determine the size of the rectangle $Z = 0.1 \times \max(\text{median}_i(R_i), \text{median}_i(C_i))$.
 - 5: Create the initial matrix F_1 and the final matrix F^* .
 - 6: **for** each image **do**
 - 7: Determine N_F , the number of SIFT points, that should be used for image I_i .
 - 8: **for** each SIFT point **do**
 - 9: Extract features f_1, f_2, \dots, f_{N_T} from the $Z \times Z$ rectangle around each SIFT point.
 - 10: Append the features as a new row to the initial matrix F_1 , which becomes of size $N_F \times N_T$.
 - 11: **end for**
 - 12: Calculate S_T different statistics from F_1 and assigned in F_2 .
 - 13: Append F_2 of the current image of size $S_T \times N_T$ to the feature matrix F_3 (the feature matrix F_3 becomes of size $L \times N_T$, $L = S_T * N_I$)
 - 14: **end for**
 - 15: Remove very similar features from F_3 (e.g., at least 99% correlated). F_4 is a reduced matrix of F_3 of size $L \times N_{T_1}$, $N_{T_1} \leq N_T$.
 - 16: Determine the number of features by discarding similar ones from F_4 (e.g., at least 90% correlated). F_C is a feature matrix generated from F_4 of size $L \times N_{T_2}$, $N_{T_2} \leq N_{T_1}$.
 - 17: Use k different unsupervised feature selection methods to generate k different feature matrices in addition to F_C : F_P , F_M , F_F , F_G , and F_L . All of these matrices are of size $L \times N_{T_2}$.
 - 18: Select any features found in at least half of the matrices to form F_5 of size $L \times N_{T_3}$, $N_{T_3} \leq N_{T_2}$.
 - 19: Generate a final feature matrix F^* from F_5 by removing similar features (e.g., at least 90% correlated). F^* is of size $L \times N_L$, $N_L \leq N_{T_3}$.
-

The next step is to calculate S_T different statistical measures from F_1 (e.g., $S_T = 8$: mean, median, mode, standard

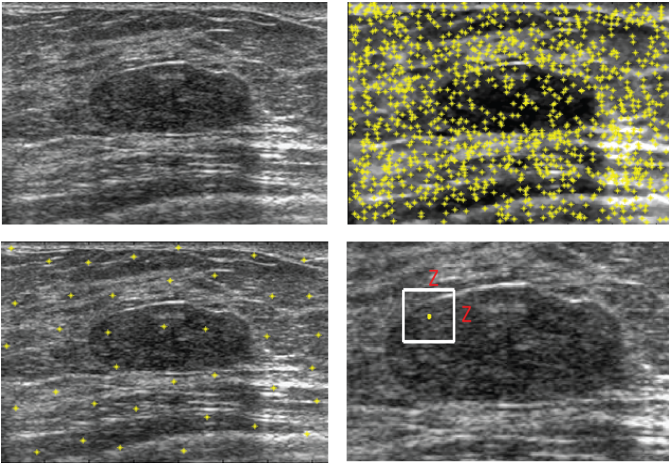


Fig. 1. Feature extraction process (from top left to bottom right): original image, seed points detected by SIFT, selected seed points via sorting the descriptor, calculating features around each selected seed point.

deviation, co-variance, range, minimum, and maximum). The resulting matrix F_2 (size $S_T \times N_T$) is returned, in which each row represents a statistical measure (Algorithm 2, *CSF*). F_2 is then appended to the feature matrix F_3 (Algorithm 2). After all images are processed, the feature matrix F_3 is formed from the features of all images, with each image being represented by S_T rows.

In the last step, the final set of features that should be used in the current image category are selected from F_3 . This process starts with the removal of very similar features in F_3 based on the calculation of the correlations between all features. Hence, if two features are highly correlated, e.g. with a correlation coefficient of at least 99%, then one is kept and the other is discarded. The output of this process is a matrix F_4 (Algorithm 2).

For any unsupervised feature selection technique, the number of features N_{T_2} that should be returned must be established in advance. A correlation with a threshold of 90% is used in order to determine the number of features that should be returned from F_4 (Algorithm 2). Following this process, F_C is the resulting feature matrix. In addition to F_C , five different unsupervised feature selection methods are also used for feature selection. The matrix F_4 and the variable N_{T_2} are passed to the methods, and each method returns a different matrix with its selected features. The resulting matrices are F_G [7], F_L [8], F_F [16], F_P [28], and F_M [4] (Algorithm 2). For all features in the six matrices, any feature extracted by at least three of the six methods are selected and appended to a matrix F_5 (Algorithm 2). The final matrix F^* is generated based on the discarding of features from F_5 that are at least 90% correlated (Algorithm 2).

B. Offline Phase

In the offline phase, the best parameters for segmenting each image are calculated through an exhaustive search and then stored in matrix T (Algorithm 3, *BSP*). The process is performed as explained in [19].

C. Training Phase

In this phase, the features selected for the training images are used for the training of the fuzzy system. A set of images are randomly selected for training (Algorithm 3). A matrix M is created and filled with the rows from F^* that belong to the training images (Algorithm 3). A matrix O is created and filled with the rows from T that belong to the training images (Algorithm 3). A pruning step is performed starting from the second training image in order to ensure that M and O do not contain similar rows (Algorithm 3). The pruned matrices M and O are used for the generation of the initial fuzzy rules (Algorithm 3). The initial fuzzy system is built through the creation of a set of rules using the Takagi-Sugeno approach to describe the in- and output matrices. Based on N_L different features from the input and one optimal parameter as the output, a set of rules is generated whereby the features are in the antecedent part and the optimal parameters are in the consequent part of the rules.

Algorithm 3 Offline and Training Phases

- 1: **Offline phase**
 - 2: Determine the parent algorithm(s) and their parameters p_1, p_2, \dots, p_k .
 - 3: Read the gold standard images G_1, G_2, \dots, G_n .
 - 4: Via exhaustive search or trial-and-error comparisons with gold standard images, determine the best segments S_1, S_2, \dots, S_n and the best parameters $p_1^*, p_2^*, \dots, p_k^*$ that generate the best segments and store them in matrix T .
 - 5: **Training phase**
 - 6: Determine the available training images I_1, I_2, \dots, I_{N_R} .
 - 7: Create two empty matrices M for input and O for output.
 - 8: **for all** N_R images **do**
 - 9: Fill matrix F_T with rows from matrix F^* that belong to the training image I_i ($F_T = F^*(I_i)$).
 - 10: Fill matrix T_R with rows from matrix T that belong to the training image I_i ($T_R = T(I_i)$).
 - 11: **if** $i=1$ **then**
 - 12: Append F_R to M , and T_R to O .
 - 13: **else**
 - 14: Pruning step: Discard rows from F_R and T_R that are similar to rows in M and O , respectively.
 - 15: Append the updated matrices F_R and T_R to M and O respectively.
 - 16: **end if**
 - 17: **end for**
 - 18: Generate fuzzy rules R_{F_1}, R_{F_2}, \dots from the input matrix M and the output matrix O (e.g., using clustering).
-

D. Online and Evolving Phase

The evolving process is performed in order to increase the capabilities of the proposed system. For each test image, a matrix F_S is filled with the rows from F^* that belong to the test image (Algorithm 4). Fuzzy inference using F_S is applied, and a parameter vector T_O is returned (size 1×8) and the final output parameter T^* is calculated (Algorithm 4). The resulting parameter is used for the segmentation of the image (Algorithm 4), and the resulting segment is stored and then displayed to the

user for review and eventual correction (Algorithm 4). The best parameter for the current image is then calculated based on the user-corrected segment and is stored in T_B (Algorithm 4). A pruning procedure is performed on F_S and T_B as described in [19], with the exception that the Euclidean distance thresholds are, in contrast to EFIS, different for different techniques. After pruning, revised versions of F_S and T_B are appended to M and O (Algorithm 4). In the final step, the current fuzzy inference system, i.e., its rule base, is regenerated using the updated matrices M and O (Algorithm 4), and the process is repeated as long as new images are available.

Algorithm 4 Online/Evolving Phase

- 1: Load the fuzzy rules R_{F_i} and the matrices M , O , and F^* .
 - 2: Load the test images I_1, I_2, \dots, I_{N_E} .
 - 3: **for** all N_E images **do**
 - 4: Fill matrix F_S with the rows from matrix F^* that belong to the test image I_i ($F_S = F^*(I_i)$).
 - 5: Perform fuzzy inference to generate output:
 $T_O = \text{FUZZY-INFERENCE}(R_{F_1}, R_{F_2}, \dots)$.
 - 6: Generate a single output T^* from T_O using the mean of T_O (μ_{T_O}), the median of T_O (M_{T_O}), the fuzzy membership (m_{T_O}) of the standard deviation of T_O (σ_{T_O}) using a Z-shaped function (zmf)
 $m_{T_O} = zmf(\sigma_{T_O}, [(\mu_{T_O} * 0.10) (\mu_{T_O} * 0.20)])$, and
 $T^* = m_{T_O} * \mu_{T_O} + (1 - m_{T_O}) * M_{T_O}$.
 - 7: Apply the parameters to segment I_i .
 - 8: Display segment S and wait for user feedback (user generates a gold standard image G by editing S)
 - 9: — ***Rule Evolution - Invisible to User*** —
 - 10: Determine the best output vector $p_1^*, p_2^*, \dots, p_k^*$ (via comparison of S with G) and store it in T_B .
 - 11: Pruning – Discard rows from F_S and T_B that are similar to rows in M and O , respectively.
 - 12: Append the matrices F_S and T_B to M and O , respectively.
 - 13: Generate fuzzy rules R_{F_i} from the updated matrices M and O (e.g., using clustering).
 - 14: **end for**
-

VI. EXPERIMENTS AND RESULTS

This section describes the experiments conducted in order to test the proposed self-configuring EFIS (SC-EFIS). To build the initial fuzzy system, for each training set, a set of randomly selected images from the data set were used for the extraction of the features along with the optimum parameters as output. This initial fuzzy system was then used to test the proposed method using the remaining images. The initial fuzzy system evolves as long as new (unseen) images are fed into the system and as long as the segmentation results produced by the algorithms are corrected by an expert user in order to generate optimal parameter values. This process drives the evolution of the fuzzy rules for segmentation. During the experimentation, the training-testing cycle was repeated 10 times. The results of ten different trials for each segmentation technique and for each parent algorithm are presented in order to validate the performance of SC-EFIS. The number of rules was monitored during the evolution process in order to acquire empirical knowledge about the convergence of the evolving process.

The experimental results using an image dataset for three different segmentation techniques (region growing, global thresholding, and statistical region merging) are presented. All experiments were performed using Matlab 64-bit.

A. Image Data

The target dataset was developed from **35 breast ultrasound scans**¹ that were segmented by an image-processing expert with extensive experience in breast lesion segmentation (the second author). The images, collected from the Web, are of different dimensions, ranging from 230×390 to 580×760 pixels (Figure 2, images resized for sake of illustration). These are the same images used to introduce EFIS originally [19].

Ultrasound images are generally difficult to segment, primarily due to the presence of speckle noise and low level of local contrast. It should be noted that the segmentation of ultrasound actually does require a complete processing chain, (including proper preprocessing and post-processing steps). However, the purpose of using these images was solely to demonstrate that the accuracy of the segmentation can be increased with the application of SC-EFIS.

B. Evaluation Measures

Considering two segments S (generated by an algorithm) and G (the gold standard image manually created by an expert), we calculate the average of the Jaccard index J (area overlap) [23]:

$$J(S, G) = \frac{|S \cap G|}{|S \cup G|}, \quad (1)$$

and its standard deviation σ_J . As well, the 95% confidence interval (CI) of the Jaccard index CI_J is calculated. Finally, we performed t-tests to validate the null hypothesis for comparing the results of a parent algorithm and its evolved version in order to establish whether any potential increase in accuracy is statistically significant. Ground-truth images G were created so that the objects of interest (i.e., lesions and tumours) could be labeled as white (1) and the background as black (0). All thresholding techniques were used consistently to label object pixels in this way as this was done in EFIS.

C. Results

To compare with EFIS, the SC-EFIS results are calculated for the same parent algorithms, namely for region growing (RG), global thresholding, and statistical region merging (SRM) are presented. The results are discussed with respect to rule evolution, visual inspection, accuracy verification using the Jaccard results.

Rule Evolution – Fig. 3 indicates the change in the number of rules during the evolving of the thresholding (THR) process. The initial number of rules increases with any incoming image and then begins to decrease as additional images become available. The same behaviour was noted for SRM and RG.

Visual Inspection – A visual inspection of Fig. 4 shows that the results produced by the proposed SC-EFIS for RG represent a substantial improvement over those obtained with

¹The images and their gold standard segments are available online: <http://tizhoosh.uwaterloo.ca/Data/>

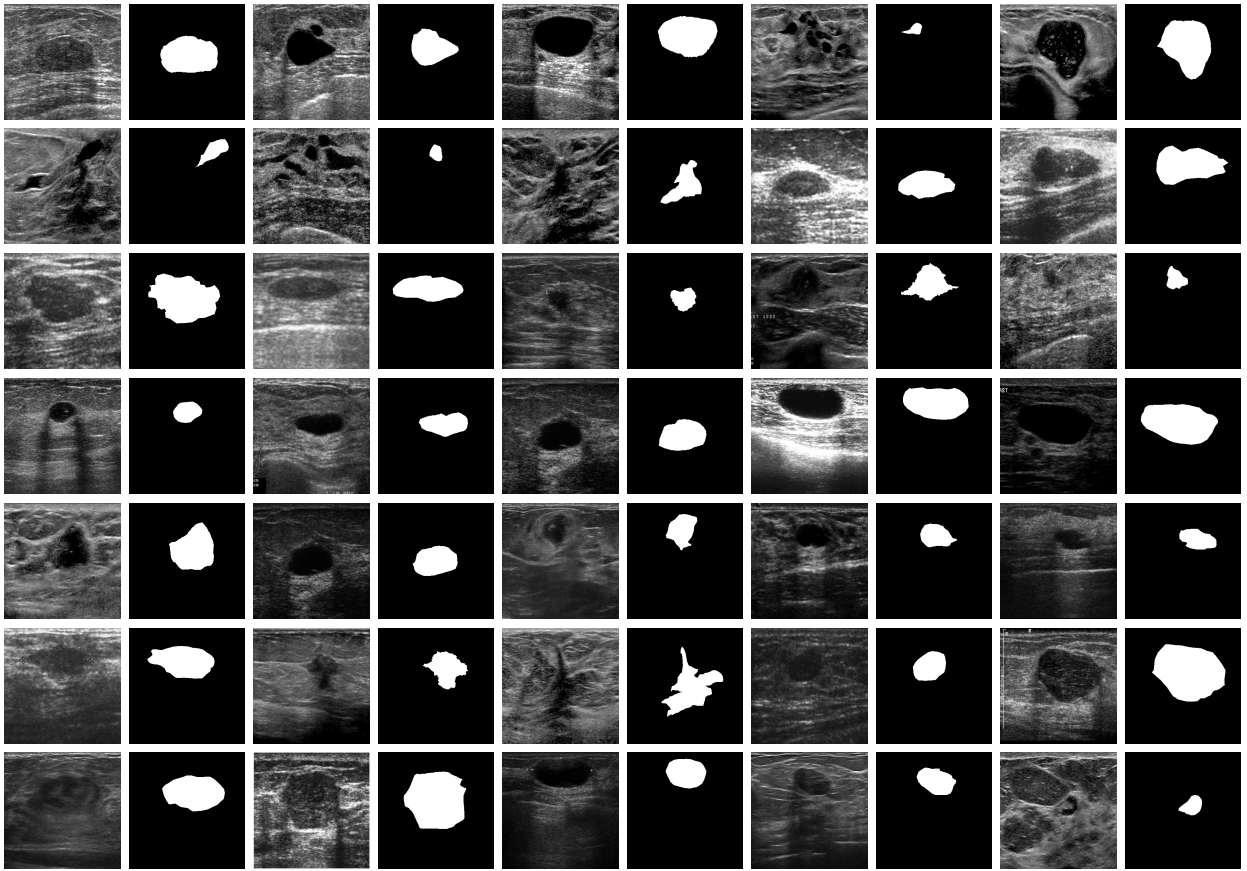


Fig. 2. Breast ultrasound scans used in our experiments. All images were segmented by an image-processing expert with extensive experience in breast lesion segmentation. Please note that some images may contain multiple ROIs. The images and their gold standard segments are available online: <http://tizhoosh.uwaterloo.ca/Data/>.

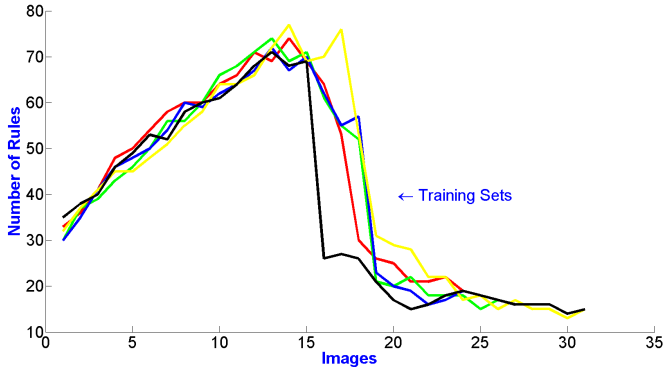


Fig. 3. Rule evolution for SC-EFIS for thresholding (THR): The number of rules increases first as more images are processed but then drops and seems to converge toward a lower number of rules. Each curve shows the number of rules for a separate trial/run.

the FRG (fuzzy RG – the initial fuzzy rules are used in order to estimate the similarity threshold). A visual inspection of Fig. 5 reveals a significant improvement in the SC-EFIS for SRM images over the SRM ones.

Accuracy Verification – Ten different trials/runs are presented for each method. Each run is an independent experiment involving different training and testing images. Fig. 6 shows the improvement in the Jaccard index of the SC-EFIS for SRM

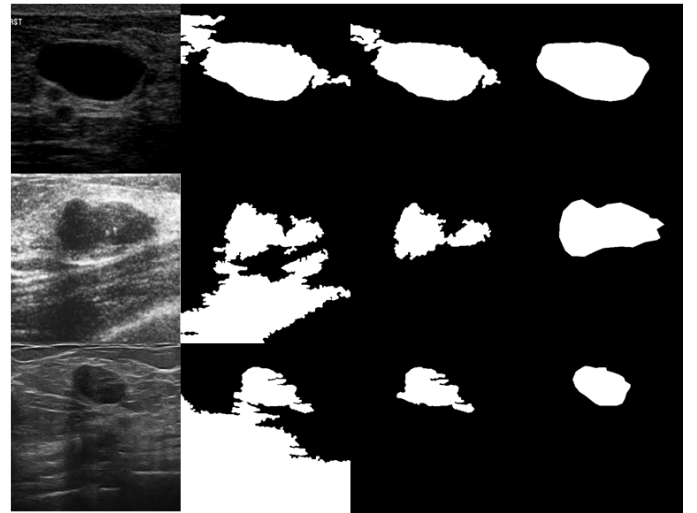


Fig. 4. Segmentation results: From left to right, the original image, FRG, SC-EFIS-RG, and the gold standard image.

and the images for SRM with a scale = 32.

Table I presents a comparison of the results for the RG technique: RG results with fuzzy inference, RG results with a similarity threshold of 0.17, RG with the best similarity threshold (0.12) for the available data (RG-B), the EFIS-RG

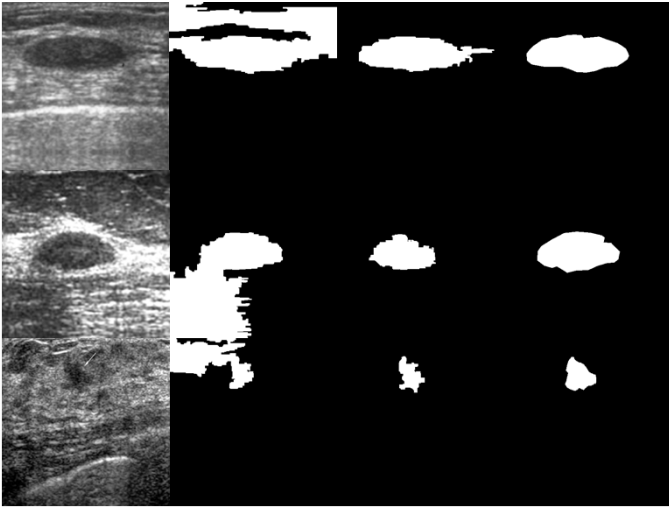


Fig. 5. segmentation results: From left to right, the original image, SRM, SC-EFIS-SRM, and the gold standard image.

technique, and the SC-EFIS-RG. The best similarity threshold, determined only for experimental purposes, is found via exhaustive search that is impractical in real world applications. It can be seen that the results achieved with SC-EFIS are better than EFIS results in eight of ten experiments.

Table II presents a comparison of the results for the global thresholding with a static (non-evolving) fuzzy system (THR) technique: the results for THR, EFIS-THR, and SC-EFIS-THR. It is clear that the SC-EFIS results surpass the EFIS ones in six of ten experiments. However, EFIS produces better results in two experiments and equivalent results in other two.

Table III presents a comparison of the results for the SRM technique: results for SRM using fuzzy inference FSRM, results for SRM with a scale = 32 (SRM), results for SRM with the best scale (64) for the available images (SRM-B) determined via exhaustive search, EFIS-SRM results, and SC-EFIS-SRM results. It can be seen that the results produced by SC-EFIS are superior to the EFIS results in five experiments, inferior in four experiments, and equivalent for the remaining experiments. Of course, both EFIS and SC-EFIS do perform better than the parent algorithm.

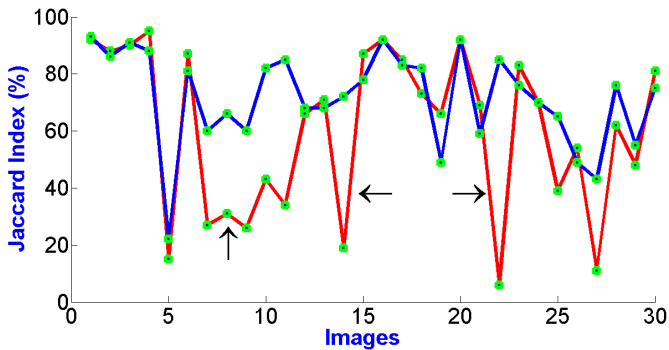


Fig. 6. Comparison of the Jaccard accuracy obtained with SC-EFIS-SRM (blue) and with SRM (red); arrows point to significant gaps.

In general, SC-EFIS is competitive with and can even surpass EFIS with respect to the three segmentation techniques,

while offering a higher level of automation.

Switching/Fusion of Results – On the other hand, the switch/fusion technique [19] was re-examined for use with SC-EFIS. Table IV presents the results of switching and fusion for the same three methods, namely Niblack, SRM (scale=32), and RG (similarity = 0.17) using EFIS (EFIS-S and EFIS-F) and using SC-EFIS (SC-EFIS-S and SC-EFIS-F). It is clear that the outcomes of EFIS and SC-EFIS are comparable. In addition, the results with EFIS-S and SC-EFIS-S surpass those for SRM, which represents the best method.

TABLE II. SAMPLE RESULTS FOR GLOBAL THRESHOLDING: FUZZY THRESHOLDING (THR), EFIS-THR, AND SC-EFIS-THR. THE NULL HYPOTHESIS WAS REJECTED IN 9/10 RUNS.

Training	Method	J	σ_J	CI_J
1st run	THR	58%	24%	49%-67%
	EFIS-THR	62%	25%	53%-71%
	SC-EFIS-THR	63%	23%	54%-72%
2nd run	THR	48%	33%	35%-60%
	EFIS-THR	61%	24%	52%-70%
	SC-EFIS-THR	61%	28%	51%-72%
3rd run	THR	43%	32%	31%-55%
	EFIS-THR	63%	25%	54%-73%
	SC-EFIS-THR	63%	26%	53%-72%
4th run	THR	23%	23%	14%-32%
	EFIS-THR	63%	22%	55%-71%
	SC-EFIS-THR	66%	21%	58%-74%
5th run	THR	54%	26%	44%-64%
	EFIS-THR	62%	24%	53%-71%
	SC-EFIS-THR	63%	25%	54%-73%
6th run	THR	55%	30%	44%-66%
	EFIS-THR	63%	23%	55%-72%
	SC-EFIS-THR	64%	23%	55%-72%
7th run	THR	38%	27%	28%-48%
	EFIS-THR	60%	24%	51%-69%
	SC-EFIS-THR	59%	26%	49%-69%
8th run	THR	52%	24%	43%-62%
	EFIS-THR	62%	21%	54%-70%
	SC-EFIS-THR	63%	21%	55%-70%
9th run	THR	39%	31%	28%-51%
	EFIS-THR	63%	23%	54%-73%
	SC-EFIS-THR	65%	21%	57%-73%
10th run	THR	44%	25%	34%-53%
	EFIS-THR	58%	26%	48%-68%
	SC-EFIS-THR	57%	26%	47%-67%

Table V enables a comparison of EFIS and SC-EFIS results for global thresholding with different global and local thresholding techniques. The data listed are taken from three experiments selected from Table II. It is clear that, in the three experiments, EFIS and SC-EFIS provide outcomes that are more accurate than those produced with the non-evolutionary thresholding techniques.

VII. CONCLUSIONS

Most image segmentation techniques involve multiple parameters that must be tuned in order to achieve maximum segmentation accuracy. Evolving fuzzy image segmentation

TABLE I. SAMPLE RESULTS FOR FUZZY REGION GROWING (FRG), RG WITH A SIMILARITY THRESHOLD (0.17), RG-B WITH THE BEST SIMILARITY THRESHOLD (0.12) (DETERMINED VIA EXHAUSTIVE SEARCH), EFIS-RG, AND SC-EFIS-RG. THE NULL HYPOTHESIS WAS REJECTED IN 10/10 RUNS.

Training	Metrics	FRG	RG	RG-B	EFIS-RG	SC-EFIS-RG
1st run	J	63%	54%	69%	68%	67%
	σ_J	26%	30%	21%	21%	23%
	CI_J	53%-73%	43%-65%	62%-77%	60%-76%	58%-75%
2nd run	J	37%	52%	69%	63%	66%
	σ_J	35%	31%	19%	24%	22%
	CI_J	24%-50%	41%-64%	62%-76%	54%-72%	57%-74%
3rd run	J	43%	54%	70%	65%	68%
	σ_J	31%	30%	21%	25%	21%
	CI_J	31%-54%	43%-65%	63%-78%	55%-74%	61%-76%
4th run	J	33%	54%	71%	64%	66%
	σ_J	33%	31%	20%	23%	24%
	CI_J	21%-46%	42%-65%	63%-78%	56%-73%	57%-74%
5th run	J	46%	54%	71%	66%	67%
	σ_J	32%	29%	17%	21%	20%
	CI_J	34%-58%	43%-65%	64%-77%	58%-74%	60%-74%
6th run	J	46%	52%	69%	64%	62%
	σ_J	31%	30%	20%	23%	24%
	CI_J	35%-58%	41%-63%	61%-76%	55%-73%	53%-71%
7th run	J	61%	57%	70%	67%	68%
	σ_J	28%	29%	21%	24%	23%
	CI_J	51%-71%	46%-68%	62%-78%	58%-75%	59%-76%
8th run	J	56%	53%	70%	64%	67%
	σ_J	30%	30%	20%	25%	23%
	CI_J	45%-67%	42%-64%	62%-78%	55%-73%	59%-75%
9th run	J	37%	53%	70%	64%	66%
	σ_J	29%	31%	20%	25%	23%
	CI_J	26%-48%	41%-64%	63%-78%	55%-73%	58%-75%
10th run	J	57%	57%	71%	66%	69%
	σ_J	29%	29%	18%	23%	21%
	CI_J	46%-68%	46%-68%	64%-78%	58%-75%	61%-77%

(EFIS) has been recently proposed to provide evolving and user-oriented adjustment for medical image segmentation. EFIS is a generic segmentation scheme that relies on user feedback in order to improve the quality of segmentation. Its evolving nature makes this approach attractive for applications that incorporate high-quality user feedback, such as in medical image analysis. However, EFIS entails some limitations, such as parameters that must be selected prior to the running of the algorithm and the lack of an automated feature selection component. These drawbacks restrict the use of EFIS to specific categories of images. An improved version of EFIS, called self-configuring EFIS (SC-EFIS) was proposed in this paper. SC-EFIS is a generic image segmentation scheme that does not require setting of some parameters, such as number of features or detecting a region of interest. SC-EFIS operates with the data available and extracts major parameters necessary for its operation from those data. A comparison of the SC-EFIS results with those obtained with EFIS demonstrates the comparable accuracy of both schemes with SC-EFIS offering a much higher level of automation.

REFERENCES

- [1] E. ARVACHEH AND H. TIZHOOSH, *Pattern analysis using zernike moments*, in Proceedings of the IEEE Instrumentation and Measurement Technology Conference (IMTC 2005), vol. 2, 2005, pp. 1574–1578.
- [2] F. BELLAL, H. ELGHAZEL, AND A. AUSSEM, *A semi supervised feature ranking method with ensemble learning*, Pattern Recognition Letters, (2012).
- [3] J. CADENAS, M. CARMEN GARRIDO, AND R. MARTNEZ, *Feature subset selection filter-wrapper based on low quality data*, Expert Systems with Applications, (2013).
- [4] D. CAI, C. ZHANG, AND X. HE, *Unsupervised feature selection for multi-cluster data*, in Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2010, pp. 333–342.
- [5] R. CAI, Z. ZHANG, AND Z. HAO, *Bassum: A bayesian semi-supervised method for classification feature selection*, Pattern Recognition, 44 (2011), pp. 811–820.
- [6] G. DOQUIRE AND M. VERLEYSEN, *A graph laplacian based approach to semi-supervised feature selection for regression problems*, Neurocomputing, (2013).
- [7] A. K. FARAHAT, A. GHODSI, AND M. S. KAMEL, *Efficient greedy feature selection for unsupervised learning*, Knowledge and Information Systems, (2012), pp. 1–26.
- [8] X. HE, D. CAI, AND P. NIYOGI, *Laplacian score for feature selection*, Advances in Neural Information Processing Systems, 18 (2006), p. 507.
- [9] L. HUANG AND M. WANG, *Image thresholding by minimizing the measure of fuzziness*, Pattern Recognition, 28 (1995), pp. 41–51.
- [10] M. KALAKECH, P. BIELA, L. MACAIRE, AND D. HAMAD, *Constraint scores for semi-supervised feature selection: A comparative study*,

TABLE III. SAMPLE RESULTS FOR FUZZY STATISTICAL REGION MERGING (FSRM), SRM WITH THE DEFAULT SCALE (32), SRM-B WITH THE BEST SCALE (64) (DETERMINED VIA EXHAUSTIVE SEARCH), EFIS-SRM, AND SC-EFIS-SRM. THE NULL HYPOTHESIS WAS REJECTED IN 10/10 RUNS.

Training	Metrics	FSRM	SRM	SRM-B	EFIS-SRM	SC-EFIS-SRM
1st run	J	64%	60%	72%	71%	72%
	σ_J	24%	28%	21%	19%	17%
	CI_J	55%-73%	50%-71%	64%-79%	64%-78%	65%-78%
2nd run	J	66%	60%	68%	69%	67%
	σ_J	25%	27%	24%	22%	20%
	CI_J	57%-76%	50%-70%	59%-76%	61%-77%	60%-75%
3rd run	J	63%	61%	70%	67%	69%
	σ_J	25%	28%	22%	24%	18%
	CI_J	53%-72%	50%-71%	62%-78%	58%-76%	62%-76%
4th run	J	57%	59%	69%	71%	71%
	σ_J	29%	30%	24%	21%	19%
	CI_J	46%-67%	48%-70%	60%-78%	63%-79%	64%-78%
5th run	J	42%	59%	68%	67%	68%
	σ_J	33%	29%	24%	23%	22%
	CI_J	30%-54%	49%-70%	59%-77%	59%-76%	60%-77%
6th run	J	63%	60%	69%	69%	68%
	σ_J	26%	28%	22%	21%	20%
	CI_J	53%-73%	49%-70%	61%-77%	61%-76%	61%-76%
7th run	J	55%	61%	70%	70%	70%
	σ_J	30%	29%	23%	22%	20%
	CI_J	44%-67%	50%-72%	62%-79%	62%-79%	63%-78%
8th run	J	67%	59%	70%	68%	69%
	σ_J	19%	28%	22%	22%	20%
	CI_J	60%-74%	48%-69%	62%-78%	60%-76%	62%-76%
9th run	J	47%	59%	69%	71%	67%
	σ_J	31%	30%	24%	22%	24%
	CI_J	36%-59%	47%-70%	60%-78%	63%-79%	58%-76%
10th run	J	64%	61%	69%	68%	71%
	σ_J	28%	29%	24%	23%	19%
	CI_J	54%-74%	51%-72%	60%-78%	60%-77%	64%-78%

- Pattern Recognition Letters, 32 (2011), pp. 656–665.
- [11] J. KITTLER AND J. ILLINGWORTH, *Minimum error thresholding*, Pattern Recognition, (1986), pp. 41–47.
- [12] T. N. LAL, O. CHAPPELLE, J. WESTON, AND A. ELISSEEFF, *Embedded methods*, in Feature Extraction, Springer, 2006, pp. 137–165.
- [13] D. LOWE, *Object recognition from local scale-invariant features*, in Proceeding of the IEEE International Conference on Computer Vision, vol. 2, 1999, pp. 1150–1157.
- [14] ———, *Distinctive image features from scale-invariant keypoints*, International journal of computer vision, 60 (2004), pp. 91–110.
- [15] J. MARTÍNEZ SOTOCÁ AND F. PLA, *Supervised feature selection by clustering using conditional mutual information-based distances*, Pattern Recognition, 43 (2010), pp. 2068–2081.
- [16] P. MITRA, C. MURTHY, AND S. K. PAL, *Unsupervised feature selection using feature similarity*, IEEE transactions on pattern analysis and machine intelligence, 24 (2002), pp. 301–312.
- [17] L. C. MOLINA, L. BELANCHE, AND À. NEBOT, *Feature selection algorithms: A survey and experimental evaluation*, in IEEE International Conference on Data Mining ICDM, 2002, pp. 306–313.
- [18] W. NIBLACK, *An Introduction to Digital Image Processing*, Strandberg Publishing Company, Birkeroed, Denmark, 1986.
- [19] A. OTHMAN, H. R. TIZHOOSH, AND F. KHALVATI, *EFIS: Evolving fuzzy image segmentation*, IEEE Transactions on Fuzzy Systems, 22 (2014), pp. 72–82.
- [20] Y. SAEYS, I. INZA, AND P. LARRAÑAGA, *A review of feature selection techniques in bioinformatics*, Bioinformatics, 23 (2007), pp. 2507–2517.
- [21] N. SÁNCHEZ-MAROÑO, A. ALONSO-BETANZOS, AND M. TOMBILLA-SANROMÁN, *Filter methods for feature selection—a comparative study*, in Intelligent Data Engineering and Automated Learning-IDEAL 2007, Springer, 2007, pp. 178–187.
- [22] L. SONG, A. SMOLA, A. GRETTON, K. M. BORGWARDT, AND J. BEDO, *Supervised feature selection via dependence estimation*, in Proceedings of the 24th international conference on Machine learning, ACM, 2007, pp. 823–830.
- [23] K. V. TAN P.-N., STEINBACH M., *Introduction to Data Mining*, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2005.
- [24] H. R. TIZHOOSH, *Image thresholding using type II fuzzy sets*, Pattern Recognition, 38 (2005), pp. 2363–2372.
- [25] ———, *Type II fuzzy image segmentation*, in Fuzzy Sets and Their Extensions: Representation, Aggregation and Models Studies in Fuzziness and Soft Computing, vol. 220, 2008, pp. 607–619.
- [26] S. WARFIELD, *Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation*, IEEE Transactions on Medical Imaging, 23 (2004), pp. 903–921.
- [27] Z. ZHAO AND H. LIU, *Semi-supervised feature selection via spectral analysis*, in Proceedings of the 7th SIAM International Conference on Data Mining, Minneapolis, MN, 2007, pp. 1151–1158.
- [28] ———, *Spectral feature selection for supervised and unsupervised learning*, in Proceedings of the 24th international conference on Machine learning, ACM, 2007, pp. 1151–1157.

TABLE IV. ACCURACY OF SWITCHING AND FUSION FOR THREE METHODS: NIBLACK, SRM, AND RG USING EFIS AND SC-EFIS: EACH DATASET HAD 30 IMAGES FOR TRAINING AND 5 IMAGES FOR TESTING.

Dataset	Niblack	SRM	RG	EFIS-S	EFIS-F	SC-EFIS-S	SC-EFIS-F
1	76%	68%	50%	77%	77%	76%	65%
2	52%	55%	48%	53%	53%	62%	52%
3	77%	74%	72%	80%	72%	80%	81%
4	74%	57%	55%	55%	56%	65%	66%
5	43%	33%	33%	36%	36%	34%	28%
6	59%	59%	62%	62%	61%	61%	57%
7	55%	82%	80%	81%	78%	62%	78%
8	62%	62%	58%	66%	65%	63%	58%
9	68%	64%	63%	76%	70%	73%	69%
10	59%	90%	89%	79%	79%	76%	90%
m	62.3%	64.5%	61.0%	66.5%	64.6%	64.9%	64.3%
σ	11%	16%	16%	15%	13%	14%	17%

TABLE V. COMPARISON OF EFIS, SC-EFIS, AND 4 OTHER GLOBAL THRESHOLDING TECHNIQUE AS WELL AS ONE LOCAL THRESHOLDING METHOD ([24], [25], [18], [11], [9]): AVERAGE AND STANDARD DEVIATION OF THE JACCARD INDEX $J \pm \sigma_J$ AND 95% CONFIDENCE INTERVAL CI_J . THE MAA INDICATES THE MAXIMUM ACHIEVABLE ACCURACY DETERMINED VIA EXHAUSTIVE SEARCH AND THROUGH COMPARISON WITH GOLD STANDARD IMAGES; NO GLOBAL THRESHOLDING METHOD CAN ACHIEVE HIGHER ACCURACIES THAN MAA.

Run	Method	$J \pm \sigma_J$	CI_J
1	MAA	79%±12%	[75% 84%]
	EFIS-THR	62%±25%	[53% 71%]
	SC-EFIS-THR	63%±23%	[54% 72%]
	Niblack (local)	56%±24%	[47% 65%]
	Huang	45%±27%	[35% 55%]
	Kittler	39%±32%	[27% 51%]
	Tizhoosh	35%±32%	[23% 47%]
	Otsu	28%±25%	[18% 37%]
2	MAA	79%±11%	[75% 83%]
	EFIS-THR	60%±24%	[51% 69%]
	SC-EFIS-THR	59%±26%	[49% 69%]
	Niblack (local)	57%±25%	[48% 66%]
	Huang	44%±29%	[34% 55%]
	Kittler	41%±31%	[29% 52%]
	Tizhoosh	38%±32%	[26% 50%]
	Otsu	29%±25%	[19% 38%]
3	MAA	79%±12%	[74% 83%]
	EFIS-THR	63%±23%	[54% 71%]
	SC-EFIS-THR	65%±21%	[57% 73%]
	Niblack (local)	59%±24%	[49% 68%]
	Huang	46%±27%	[35% 56%]
	Kittler	41%±33%	[29% 53%]
	Tizhoosh	35%±33%	[23% 48%]
	Otsu	28%±23%	[20% 37%]